**SCHOOL OF BUSINESS INFORMATICS AND MATHEMATICS**
**Chair of Data Science**
**Prof. Dr. Heiko Paulheim**

UNIVERSITÄT MANNHEIM

Universität Mannheim · Chair of Data Science
B 6, 26 · 68131 Mannheim

To the
SWSA Distinguished Dissertation
Award Committee

**Prof. Dr. Heiko Paulheim**
University of Mannheim
D-68131 Mannheim
Tel.: +49 621 181 2646
Mail: heiko@informatik.uni-mannheim.de

**Visitor address:**
B 6, 26
D-68131 Mannheim

http://dws.informatik.uni-mannheim.de/

**Mannheim, 26. Juni 2018**

Letter of Support for Nomination of Kalpa Gunaratna's Dissertation "Semantics-based Summarization of Entities in Knowledge Graphs"

Dear Members of the Award Committee:

As a researcher dedicated to knowledge graphs and their usage, I have followed Kalpa Gunaratna's work throughout the past years, published in top-tier venues such as AAAI, ESWC, and IJCAI. It is my great pleasure to support the nomination of Kalpa Gunaratna's dissertation for the SWSA Distinguished Dissertation Award. Let me briefly elaborate more below to support the claims.

Knowledge graphs have become extremely important with the improvements in large scale processing and machine learning techniques. With ever-increasing size of the knowledge graphs in both the number of entities and the number of facts per entity, efficient presentation methods for both human and machine users are imperative. This is exactly the theme of Kalpa's dissertation where solutions are proposed for single and multi-entity summarization. The importance of efficient and effective presentation techniques of knowledge graphs have real-world implications, as observed, e.g., for the Google Knowledge Graph. In fact, most major search engines are backed by a comprehensive knowledge graph that interacts with the users through entity summaries combined with the search results.

The first of the three main contributions of this dissertation is the diversity-aware summarization of entity descriptions (the FACES approach published in AAAI 2015). What is novel in this approach compared to the state-of-the-art approaches is its semantic grouping of facts (i.e., RDF triples of entity descriptions). The clustering approach adapted for this is incremental and the facts in a group agree on abstraction making them more semantically related. These groupings are used to model the diversity of facts and hence, the selection of facts from various groups made the summaries to have high quality diversity compared to comparable systems' primitive lexical-based diversity. In addition, I believe the incremental grouping approach adapted in this work could be used in other Semantic Web applications needing dynamic processing capabilities (like stream processing).

The second contribution of Kalpa's dissertation has a close connection to one of my recent research efforts, i.e., type prediction in knowledge graphs such as DBpedia. In our ISWC 2013 paper on "Type inference on noisy RDF data" and subsequent publications, the focus was on inferring types for entities where the types are missing. Types are very important for semantic understanding of data and the approach proposed in our 2013 paper was successfully integrated into the DBpedia knowledge graph extraction framework to improve the type coverage for entities. Similarly, Kalpa's extension to the FACES summarization approach (i.e., FACES-E, published at ESWC 2016) is to make use of

literals in datatype properties by computing semantic types for them. While every type inference and computation method in knowledge graphs is focused on entities, Kalpa's effort to make literals more intelligent stands out for its novelty in both application domain and methodology. I was there at the ESWC 2016 conference when this was presented and there were interesting discussions on this line of research during the presentation and afterwards. Even though Kalpa's application of type computation of literals was to improve entity summarization, I strongly believe this starting point of making literals more usable is a crucial step of making data more semantic and interoperable, according to the original Semantic Web vision and later formulations (e.g., Google's claim for "things, not strings").

Kalpa's third contribution of multi-entity summarization (the REMES system published at IJCAI 2017) is interesting and challenging. It is interesting because more practical usage is there for knowledge exploration applications where multiple entities are present (e.g., reading comprehension applications). The challenging aspect is that it is computationally expensive to process multiple entities concurrently to maximize an objective, which is to select more related facts for multiple entities. I strongly believe that this initiative for more complex processing tasks for entity summarization is only the beginning of more interesting research contributions to follow, which could be brought to the extreme of providing personalized and/or contextualized versions of entire knowledge graphs.

In summary, in my opinion, this dissertation covered a timely and interesting topic in the Semantic Web area, covering an extremely useful subject area of knowledge graphs. The results of the main contributions of this dissertation are published in well-known top-tier extensively peer reviewed conferences including AAAI, ESWC, and IJCAI, which makes for an impressive top-tier publication record which is rather unusual and remarkable for a PhD student. Furthermore, Kalpa did not only pursue his research, but also became an active and acclaimed member of the community. More specifically, topics covered in his dissertation materialized into a workshop series at ESWC, where I was also involved in the 2015 edition as a co-organizer. Therefore, I can speak of this work's importance by also going beyond simple publication cycle and building and supporting a community around the focus of the dissertation research. This is also an exciting validation of Kalpa's work on timeliness and importance to the community. Kalpa's advisor, Prof. Dr. Amit Sheth, who for over a decade has been one of the most accomplished researchers in the Semantic Web field, also deserves my recommendation for directing this comprehensive work. Accordingly, I strongly recommend his dissertation for the SWSA Distinguished Dissertation Award.

Prof. Dr. Heiko Paulheim
Chair of Data Science
University of Mannheim